used as the trained patterns to be selected. When the trained patterns of category 3 are selected in step S405, average values 313 of utterance by training speakers in category 3 and covariance values 323 of utterances by the training speakers in category 3 are used as the trained patterns to be selected. Now, pattern by-characteristic selection unit 204 finishes its process.

Speaker adaptation processor 205 distorts a spectral frequency on an LPC cepstral coefficient vector by Oppenheim method equation (2) using a first utterance part of the vector of the input speech sound, where the vector has been already calculated by acoustic analysis unit 202. The Oppenheim method is also disclosed in Oppenheim, A.V. and Johnson, D.H. "Discrete Representation of Signals," Proc. IEEE 60 (6): 681-691 (1972).

A distance measure of the utterance is calculated between the LPC cepstral coefficient vector, of which spectral frequency has been distorted, and a pattern arrangement corresponding to the vocabulary indicating the controlled device. The pattern arrangement has been generated using the trained patterns determined by pattern selection unit 204. In other words, LPC cepstral coefficient vector $\vec{X}^{\alpha}$ is obtained by distorting input LPC cepstral coefficient vector $\vec{X}$ through a filter shown by equation (2) using a frequency distortion coefficient α. The frequency distortion coefficient providing the most similar distance of all vector $\vec{X}^{\alpha}$ is determined according to equation (3) in relation to LPC cepstral coefficient vector $\vec{X}^{\alpha}$.

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \qquad (2)$$

where;

α is a vocal tract length normalization coefficient (frequency distortion coefficient).

$$\hat{\alpha} = \arg \max_{\alpha} P(\vec{X}^{\alpha} \mid \alpha, \theta) \qquad (3)$$

where;

P is a probability (similarity),

α is a vocal tract length normalization coefficient (frequency distortion coefficient),

$\vec{X}$ is an LPC cepstral coefficient vector, and

θ is a trained pattern.

A process by speaker adaptation processor 205 will be hereinafter described using a flow chart shown in Fig. 5.

Three initial values ($\alpha_{def} - \Delta\alpha_1$, α_{def}, $\alpha_{def} + \Delta\alpha_1$) of distortion coefficients of spectral frequency of a calculated object are firstly set (step S501). Preferably, α_{def} is 0.20 to 0.50 and $\Delta\alpha_1$ is 0.005 to 0.100 when a sampling frequency of speech sounds is 10kHz, and the present embodiment employs $\alpha_{def} = 0.35$ and $\Delta\alpha_1 = 0.02$.

Speaker adaptation processor 205 then calculates three sets of LPC cepstral coefficient vectors using spectral frequency distortioncalculation. (step S502). In thiscalculation, the processor 205 passes the first utterance part of the LPC cepstral coefficient vector of user's utterance through the following filter to distort the spectrum on the LPC cepstral coefficient vector (hereinafter called a spectral frequency distortioncalculation.). The. filter is represented by equation (2) using the spectral frequency distortion coefficients set in step S501. The LPC cepstral coefficients of user's utterance have been already obtained by acoustic analysis unit 202.

Next, speaker adaptation processor 205 stores the trained patterns that are determined by pattern selection unit 204 and the recognition result of the vocabulary indicating a controlled device (step S503).

Next, processor 205 calculates distances between three sets of LPC cepstral coefficient vectors determined in step S502 and a pattern arrangement formed using the trained patterns that are determined in step S503, based on the recognition result obtained in step S503 (step S504). When the device selection word determined by pattern selection unit 204 is “Television” and the trained pattern belongs to category 2, the simplified Mahalanobis' distance L is described every LPC cepstral coefficient as follows;.

$$L_{51k} = \sum_{i,time} B_{5k_i} - 2\,\vec{A}^{t}_{5k_i} \cdot \vec{X}_1$$

where;

$$\vec{A}_{5k} = \vec{W}^{-1} \cdot \vec{\mu}_{5k} - \vec{W}^{-1} \cdot \vec{\mu}_x$$

$$B_{5k} = \vec{\mu}_{5k}^{t} \cdot \vec{W}^{-1} \cdot \vec{\mu}_{5k} - \vec{\mu}_x^{t} \cdot \vec{W}^{-1} \cdot \vec{\mu}_x,$$

$\vec{\mu}_{5k}$ is an average value of LPC cepstral coefficient vectors of state (k) (phoneme order or time sequence) of an arrangement "Television" of speech sound elements for category 2,

$\vec{\mu}_x$ is an average value of LPC cepstral coefficient vectors of all utterances by training speakers in category 2,

W is a covariance value of LPC cepstral coefficient vectors of all utterances by the training speakers in category 2, and

$\vec{X}_1$ is an LPC cepstral coefficient vector when the spectral frequency distortion coefficient is 0.33..

$$L_{52k} = \sum_{i,time} B_{5k_i} - 2\vec{A}_{5k_i}^{t} \cdot \vec{X}_2$$

where;

$\vec{X}_2$ is an LPC cepstral coefficient vector when the spectral frequency distortion coefficient is 0.35.

$$L_{53k} = \sum_{i,time} B_{5k_i} - 2\vec{A}_{5k_i}^{t} \cdot \vec{X}_3$$

where;

$\vec{X}_3$ is an LPC cepstral coefficient vector when the spectral frequency distortion coefficient is 0.37.

Next, speaker adaptation processor 205 discriminates and determines a spectral frequency distortion coefficient when the most similar, namely the nearest, distance is obtained among the distances $L_{51k}, L_{52k}, L_{53k}$ obtained